

Samuel Pedro Beserra Dantas Lélis
Orientador: Romuere Rodrigues Veloso e Silva

**Aplicação de ciência de dados para a tomada
de decisão em acasalamento de bovinos e
melhoria genética**

Picos - PI
03 de março de 2023

Samuel Pedro Beserra Dantas Lélis
Orientador: Romuere Rodrigues Veloso e Silva

Aplicação de ciência de dados para a tomada de decisão em acasalamento de bovinos e melhoria genética

Monografia submetida ao Curso de Bacharelado em Sistemas de Informação como requisito parcial para obtenção de grau de Bacharel em Sistemas de Informação. Orientador: Prof. Dr. Romuere Rodrigues Veloso e Silva.

Universidade Federal do Piauí
Campus Senador Helvídio Nunes de Barros
Bacharelado em Sistemas de Informação

Picos - PI
03 de março de 2023

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca José Albano de Macêdo

L541a Lélis, Samuel Pedro Beserra Dantas

Aplicação de ciência de dados para a tomada de decisão em
acasalamento de bovinos e melhoria genética [recurso eletrônico] /
Samuel Pedro Beserra Dantas Lélis – 2023

40 f.

1 Arquivo em PDF

Indexado no catálogo *online* da biblioteca José Albano de Macêdo-
CSHNB

Aberto a pesquisadores, com restrições da Biblioteca

Trabalho de Conclusão de Curso (Graduação) – Universidade Federal
do Piauí, Bacharelado em Sistemas de Informação, Picos, 2023.

“Orientador: Dr. Romuere Rodrigues Veloso e Silva.”

1. Pecuária - bovinos. 2. Matriz de parentesco. 3. Endogamia. 4.
Python. I. Silva, Romuere Rodrigues Veloso e. II. Título.

CDD 636.2

APLICAÇÃO DE CIÊNCIA DE DADOS PARA A TOMADA DE DECISÃO EM
CRUZAMENTOS DE GADO BOVINO E MELHORIA GENÉTICA

SAMUEL PEDRO BESERRA DANTAS LÉLIS

Monografia _____ Aprovada _____ como exigência parcial para obtenção do
grau de Bacharel em Sistemas de Informação.

Data de Aprovação

Picos – PI, 14 de março de 2023


Prof. Romuere Rodrigues Veloso e Silva

Documento assinado digitalmente

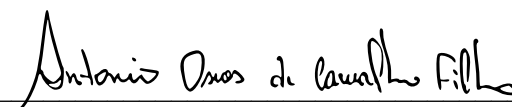


JOSE LINDENBERG ROCHA SARMENTO

Data: 17/03/2023 15:57:48-0300

Verifique em <https://validar.it.gov.br>

Prof. José Lindenberg


Prof. Antonio Oseas

Agradecimentos

Agradeço ao Sr. Geraldo Lélis, mais conhecido como meu pai, a Sra. Iva Beserra, mais conhecida como minha mãe e que desde sempre estão ao meu lado apoiando incondicionalmente minhas escolhas. Afirmo sem nenhuma dúvida que sem vocês eu não teria êxito nesta caminhada, muito obrigado por todos os ensinamentos, suporte e apoio sempre. Agradeço também a todos meus familiares que me deram apoio nessa trajetória.

Agradeço a meu orientador Prof. Dr. Romuere Rodrigues Veloso e Silva, pela sabedoria com que me orientou nesta trajetória, pelas exigências, críticas e paciência em toda essa jornada. E a todos os professores do curso por darem o melhor de si para repassar seus conhecimentos tão importantes na minha formação profissional.

Gostaria de agradecer também à minha namorada Tayla Nascimento pelo enorme apoio na conclusão desse desafio, especialmente nos momentos mais difíceis da jornada.

Por último, mas não menos importante, gostaria de agradecer a meus amigos pelo apoio e pelos momentos de descontração, tanto aos que influenciaram minha decisão de iniciar a jornada, Matheus Torres, Lucas Cardoso, Gabriel Carvalho e João Victor Carvalho, quanto aos que fiz ao longo do caminho, Lucas Bezerra, Gabriell Oliveira, Marcos Paulo, Lucas Sousa, Carlos Victor, Elievelto Edimar, Gabriel Holanda, Jederilson Sousa, Lucas Vinícius, Samuel Oliveira e Thaliane Gomes.

Sem dúvidas esses parágrafos não contêm todas as pessoas que passaram por esta fase tão importante da minha vida. Portanto, peço desculpas àqueles que não estiverem presentes dentre minhas palavras, mas tenham certeza que por mais curta que tenha sido a nossa aproximação, sempre farão parte do meu pensamento, das minhas memórias e tem minha gratidão por me ajudarem a crescer de alguma forma, de me ajudar a ser quem eu sou hoje. A todos, meu muito obrigado.

Nunca é tarde pra ser quem você poderia ter sido.

George Elliot

Resumo

O setor pecuário vem se expandindo e se consolidando cada vez mais pelo Brasil. Atualmente, o Brasil é um dos maiores exportadores de carne bovina do mundo inteiro, sendo responsável por cerca de 13,45% de toda a produção mundial em 2021. O mercado de carne bovina não só gera renda para o Brasil como é responsável por empregar cerca de 3,3 milhões de brasileiros em 2020. A qualidade e quantidade de produção de carne depende da criação e administração dos rebanhos pelos grandes pecuaristas. O controle de qualidade do rebanho deve ser rigoroso e estar sempre procurando novas técnicas que possam tornar o processo mais eficiente e eficaz. O mau gerenciamento de um rebanho pode ser ocasionado por uma série de problemas. Um dos principais problemas de gerenciamento de bovinos está na reprodução. Indivíduos do mesmo rebanho ocasionalmente irão possuir um grau de parentesco, e assim como em humanos, o gado também sofre de problemas genéticos provenientes de relações endogâmicas. Relações endogâmicas podem provocar o surgimento de efeitos deletérios nos animais, causando aumento da taxa de mortalidade dos bezerros. Por outro lado, se um criador souber gerenciar essas relações corretamente elas podem ser utilizadas para preservação de indivíduos de raça. Atualmente, existem poucos sistemas de gerenciamento de gado que podem ajudar com questões de endogamia e que sejam acessíveis e fáceis de gerenciar. Nesse contexto, esta monografia tem como objetivo fornecer algoritmos para construir as relações de parentesco e permitir maior controle de acasalamentos. Os algoritmos podem ser utilizados para calcular o índice de parentesco entre indivíduos, que será utilizado na tomada de decisão no momento da realização do cruzamento entre indivíduos de um rebanho. Os algoritmos foram desenvolvidos e integrados a um projeto maior que visa gerenciar um rebanho como um todo, que considera também outros aspectos que fogem do escopo deste trabalho, como registros de acasalamentos entre animais, nascimento da prole, registros de rebanho ao qual cada animal pertence, dentre outras. Os resultados mostram a assertividade dos algoritmos desenvolvidos. A aplicação tem grande potencial para ajudar grandes e pequenos administradores de rebanho a melhorar a eficiência da reprodução de um rebanho.

Palavras-chaves: matriz de parentesco, pecuária, aplicação, python, endogamia.

Abstract

The livestock sector has been expanding and consolidating itself more and more throughout Brazil. Currently, Brazil is one of the largest beef exporters in the world, being responsible for 13.45% of the world's production in 2021. The beef market not only generates income for Brazil, but is also responsible for employing 3.3 million Brazilians in 2020. The quality and quantity of meat production depends a lot on the creation and management of herds by large cattle ranchers. The quality control of the herd must be strict and always looking for new techniques that can make the process more efficient and effective. Mismanagement of a herd can easily be caused by a number of problems. One of the main problems in livestock management is in breeding between individuals. Individuals from the same herd will occasionally have a certain degree of relatedness to each other, and just like humans, cattle also suffer from genetic problems arising from inbreeding relationships.

Inbreeding relationships can generally lead to deleterious effects in cattle, causing an increase in the mortality rate of calves. On the other hand, if a cattle owner knows how to manage these relationships correctly, they can be used for the preservation of breed individuals. Currently, there are few livestock management systems that can help with inbreeding issues that are affordable and easy to manage. In this context, this monograph aims to provide kinship matrix algorithms. The algorithms can be used to calculate the kinship index between individuals, which will be used in decision-making when performing the crossing between individuals of a herd. The algorithms were developed and integrated into a larger project that aims to manage a herd as a whole, which also considers other aspects that are beyond the scope of this work, such as animals registration, breeding, separation in herds and other more features. The results show the assertiveness of the developed algorithms. The application has great potential to help large and small herd managers improve the efficiency of breeding a herd.

Lista de ilustrações

Figura 1 – Os 12 princípios do manifesto ágil.	16
Figura 2 – Exemplo de quadro Kanban.	17
Figura 3 – Fluxograma de desenvolvimento.	22
Figura 4 – Esquema simplificado de tabelas do banco de dados.	26
Figura 5 – Tela de mostragem do parentesco entre 2 animais.	30
Figura 6 – Tela de mostragem do parentesco médio da raça.	31
Figura 7 – Tela de mostragem do coeficiente de endogamia médio da raça.	31
Figura 8 – Tela de mostragem do coeficiente de endogamia médio da Fazenda 1.	32
Figura 9 – Tela de mostragem do coeficiente de endogamia médio da Fazenda 2.	32
Figura 10 – Tela de mostragem do coeficiente de endogamia agrupado por ano de nascimento.	33
Figura 11 – Tela de mostragem do coeficiente de endogamia agrupado por ano de nascimento da Fazenda 1.	34
Figura 12 – Tela de mostragem do coeficiente de endogamia agrupado por ano de nascimento da Fazenda 2.	35
Figura 13 – Tela de mostragem do intervalo de gerações geral, de animais reprodutores e de animais matrizes.	35

Lista de tabelas

Tabela 1 – Comparação de trabalhos a partir dos ano, objetivo, ciência de dados e matriz de parentesco.	21
Tabela 2 – Requisitos funcionais.	23
Tabela 3 – Requisitos Não Funcionais.	25
Tabela 4 – Tabela de <i>Pedigree</i> com Data de Nascimento e Rebanho.	28

Lista de abreviaturas e siglas

EUA	Estados Unidos da América
ONU	Organização das Nações Unidas
IBD	Genes Idênticos por Descendência
BLUB	<i>Best Linear Unbiased Prediction</i>
NRM	<i>Numerator Relationship Matrix</i>
REML	Método da Máxima Verossimilhança Restrita
IBS	Idênticos por Estado
GRM	<i>Genomic Relationship Matrix</i>
DBMS	Sistema de Gerenciamento de Banco de Dados
JIT	<i>Just-In-Time</i>
WIP	Limitar o Trabalho em Progresso
MTV	<i>Model Template View</i>
UML	<i>Unified Modeling Language</i>
GrP	Grau de Parentesco
GrPm	Grau de Parentesco médio
F	Coefficiente de Endogamia
F _m	Coefficiente de Endogamia médio
F _{m_r}	Coefficiente de Endogamia médio de um rebanho r
F _{m_a}	Coefficiente de Endogamia médio de um rebanho r
IG	Intervalo de Gerações
IG _m	Intervalo de Gerações dos Reprodutores
IG _f	Intervalo de Gerações das Matrizes

Sumário

1	Introdução	12
1.1	Objetivos	13
1.2	Organização do Trabalho	13
2	Referencial Teórico	14
2.1	Endogamia	14
2.2	Matriz de Parentesco	14
2.3	Ciência de Dados	15
2.4	<i>PostgreSQL</i>	15
2.5	<i>Kanban</i>	16
2.6	<i>Django</i>	17
3	Trabalhos Relacionados	19
4	Aplicação Desenvolvida	22
4.1	Metodologia	22
4.2	Requisitos Funcionais	23
4.3	Requisitos Não Funcionais	25
4.4	Modelagem do Banco de Dados	25
4.5	Métricas Calculadas	26
5	Resultados e Discussões	28
5.1	Dados Preliminares	28
5.2	Representação dos Dados	30
6	Conclusão e Trabalhos Futuros	36
	Referências	37

1 Introdução

A criação de bovinos brasileira vem aumentando nos últimos anos e tem se consolidado no cenário nacional e internacional. Nos últimos anos o Brasil se tornou um dos maiores produtores de carne de gado do mundo, alcançando cerca de 13,45% da produção mundial em 2021(CROPS. . . , [acessado em 06 de Março de 2023](#)), ficando atrás apenas dos Estados Unidos da América. Segundo dados da Associação Brasileira das Indústrias Exportadoras de Carnes, o Brasil exportou cerca de 2,3 milhões de toneladas de carne de gado em 2022(EXPORTACOES, [acessado em 06 de Março de 2023](#)).

Segundo dados da Organização das Nações Unidas (ONU) a população global alcançou a marca de 8 bilhões de pessoas em 2023 e estima-se que ela chegue a 9,7 bilhões até 2050 (POPULACAO. . . , [acessado em 06 de Março de 2023](#)). O crescimento populacional exigirá uma maior demanda por alimento, o que trará alguns novos desafios para a agricultura e pecuária, o que por consequência leva a necessidade de maneiras cada vez mais eficientes de produzir.

O controle da qualidade é um dos problemas da produção mundial de bovinos atualmente. As técnicas que são utilizadas atualmente vão desde cuidados no pasto, até formas de conservação após o abate (HOCQUETTE et al., 2007; LIU; LANARI; SCHAEFER, 1995). Um problema pouco explorado ainda na área diz respeito a perda de qualidade e problemas genéticos que podem ser ocasionados por relações endogâmicas entre animais do rebanho. A endogamia aumenta o risco do surgimento de efeitos deletérios provenientes de genes recessivos e pode afetar a capacidade de produção de todo um rebanho (MACEDO et al., 2014). Por mais que a endogamia entre indivíduos possa ser utilizada para preservação de características e raças, ela também é associada ao surgimento de doenças e condições que vão impactar o volume de produção do gado, que implica em perda financeira. O uso de novas tecnologias pode ser a chave para ajudar a controlar a frequência de reproduções endogâmicas.

O uso de *softwares* para resolver problemas vem se tornando cada vez mais comum nas mais diversas áreas do conhecimento (GORDON; HAREL, 2014). Na área de produção de gado existem diversos tipos de software que ajudam a melhorar a eficiência da produção, como *softwares* para acompanhamento de massa corpórea, identificação de indivíduos, gerenciamento de alelos recessivos, etc (FILHO et al., 2016; BELLO; ABUBAKAR, 2019; UPPERMAN et al., 2019). O problema do gerenciamento de relações endogâmicas entre indivíduos de um rebanho é pouco explorado na literatura. Os *softwares* que fazem o controle de endogamia atualmente são pouco acessíveis ou complexos. Sem o auxílio de um software os criadores ficam a mercê de realizar manualmente um trabalho extensivo de registro de gado e controle através de planilhas de *pedigree* (KUDO, 1962).

O uso de um *software* simples e acessível se faz necessário para o controle de relações

endogâmicas por produtores de gado. Neste contexto, esta monografia visa desenvolver uma ferramenta para auxiliar no controle de relações endogâmicas com o uso da técnica de matriz de parentesco. A matriz de parentesco consiste em um método que visa mapear um rebanho e atribuir valores genéticos a cada um deles baseado no método de melhor predição linear não viciada desenvolvida por [Henderson \(1976\)](#). A matriz tem como objetivo relacionar os indivíduos mapeados de forma que seja possível identificar grau de parentesco entre eles ([PEREIRA, 2004](#)). O uso da ferramenta que foi desenvolvida irá facilitar e automatizar o cálculo do grau de parentesco entre os indivíduos do gado. O sistema oferece ferramental para o gerenciamento dos animais, contendo funcionalidades de gerenciamento de cruzamentos, nascimentos, e separação por rebanhos para auxiliar os proprietários de fazenda e fornecer informações pertinentes à tomada de decisão.

1.1 Objetivos

Esta monografia tem como objetivo geral desenvolver e expor resultados de uma ferramenta para controle de relações de parentesco e endogâmicas entre indivíduos de um rebanho utilizando o método para a construção da matriz de parentesco. Os objetivos específicos deste trabalho são:

- Aplicar conhecimentos de desenvolvimento de sistemas adquiridos durante a formação para gerar um sistema intuitivo que resolva o problema proposto;
- Melhorar a forma de gerenciar relações endogâmicas e tornar mais acessível para produtores de gado;
- Apresentar estudos de caso para validação dos dados calculados pelo sistema.

1.2 Organização do Trabalho

O restante desta monografia está dividido da seguinte forma:

- O Capítulo [2](#) apresenta alguns conceitos básicos que são necessários para a total compreensão deste trabalho;
- O Capítulo [3](#) apresenta os trabalhos relacionados;
- No Capítulo [4](#) é mostrada a metodologia que foi utilizada para o desenvolvimento desta monografia;
- O Capítulo [5](#) apresenta alguns resultados e casos de uso gerados a partir da ferramenta desenvolvida neste trabalho;
- Por fim, o Capítulo [6](#) conclui esta monografia e apresenta alguns trabalhos futuros;

2 Referencial Teórico

Este capítulo descreve conceitos relevantes para o entendimento deste trabalho. As seções incluem conceitos sobre endogamia, matriz de parentesco, ciência de dados, *PostgreSQL*, *Kanban* e *Django*.

2.1 Endogamia

A endogamia é o acasalamento de animais com parentesco, que possuem algum ancestral em comum, ou seja, um acasalamento consanguíneo. A endogamia foi utilizada com técnicas de seleção para aumentar a uniformidade em algumas raças bovinas e no processo de padronização das raças a fim de fixar os biótipos. No entanto, o cruzamento consanguíneo possui muitos efeitos negativos como a perda de vigor dos animais, principalmente para características reprodutivas, reduzindo o número de animais produzidos e a economicidade da atividade. Além disso, a endogamia pode resultar no aumento da taxa de aparecimento de defeitos genéticos relacionados a genes com alelo recessivo deletério, podendo aumentar a taxa de descarte involuntário e facilitar a perda aleatória de alelos por amostragem (ROSA et al., 2013; FONTOURA, 2019).

Apesar de seus riscos, a endogamia é bastante usada por criadores de animais, para garantir a uniformidade racial e fixação de certas características em linhagens para aumentar o valor de suas criações. Contudo, é necessário um controle com monitoramento permanente, a endogamia se mostrou diretamente proporcional à perda de massa e de desempenho reprodutivo do rebanho em animais de corte e reprodutores, respectivamente (QUEIROZ; ALBUQUERQUE; LANZONI, 2000).

2.2 Matriz de Parentesco

A maioria dos métodos de seleção precisam considerar os indivíduos de uma população para obter a melhor acurácia de predição possível. Vários métodos para descobrir o parentesco foram estudados e definiram conceitos e métodos para calcular genes idênticos por descendência (IBD). O IBD pode ser usado para indicar a probabilidade de que dois alelos homólogos tenham sido herdados a partir de um mesmo ancestral. Em 1949 Henderson desenvolveu um procedimento para realizar a predição de valores genéticos que foi chamado de *Best Linear Unbiased Prediction* (BLUP) (HENDERSON, 1975). A partir das informações de *pedigree* da população é possível calcular a probabilidade de que dois alelos sejam IBD. Dessa forma, é possível calcular a probabilidade de que dois indivíduos

compartilhem o mesmo alelo proveniente de um ancestral em comum, montando a matriz de parentesco conhecida como *Numerator Relationship Matrix* (NRM) (FARAH, 2014).

A NRM permite obter componentes de variância para uma população-base e a predição de valores genéticos de indivíduos de qualquer geração através do Método da Máxima Verossimilhança Restrita (REML). Apesar da eficiência do REML, trata-se de um processo lento pois o cálculo do parentesco entre os indivíduos é realizado como uma probabilidade destes animais apresentarem genes em comum. Contudo, muitos alelos podem ser idênticos por estado (IBS), fazendo os indivíduos mais aparentados que a média da população (POWELL; VISSCHER; GODDARD, 2010). Novas tecnologias permitiram usar informações mais precisas sobre os genes IBD e IBS que podem ser compartilhados por meio de ancestrais comuns, ausentes no *pedigree*, criando uma nova matriz de parentesco genômica denominada *Genomic Relationship Matrix* (GRM). A GRM pode ser utilizada para calcular parentescos genômicos através de uma estimativa da proporção de raças que compõe os indivíduos sob avaliação genética (JULIUS et al., 2012; FARAH, 2014).

2.3 Ciência de Dados

A tecnologia se expandiu exponencialmente resultando em um aumento da geração de dados por usuários, dispositivos e sistemas, dando origem ao termo *Big Data*. Essas grandes quantidades de dados não geraram apenas preocupações de armazenamento, mas também oportunidades para análise, entendimento, modelagem e predição de diversas variáveis (SILVA et al., 2017). A ciência de dados surgiu para explorar essas oportunidades.

A exploração do *Big Data* obteve visibilidade a ponto de organizações públicas e privadas passarem a coletar informações específicas de domínio, que podem conter informações úteis sobre problemas como segurança cibernética, detecção de fraudes e marketing.

2.4 PostgreSQL

A evolução da tecnologia e fenômenos como o *Big Data* trouxeram a necessidade de sistemas para armazenamento e gerenciamentos de dados. Os bancos de dados (BD) surgiram para armazenar esses dados de forma organizada e eficiente. Contudo, a medida que a quantidade de dados armazenadas nos BDs se tornavam maiores, mais lentas as buscas se tornavam, além da complexidade para manter os BDs organizados. Os sistemas de gerenciamento de banco de dados (DBMS) surgiram com a proposta de manipular grandes bases de dados. O DBMS permite a interrogação, manutenção e análise de dados, além de fornecer interfaces com uma ampla gama de linguagens para ajudar todos os tipos de usuários. Muitas empresas produziram DBMS para o mercado, porém podem cobrar altos valores para disponibilizá-los para uso comercial (FRY; SIBLEY, 1976; RAMAKRISHNAN; GEHRKE; GEHRKE, 2003).

O *PostgreSQL* é um DBMS de código aberto. O *PostgreSQL* ganhou muita popularidade por se tratar de uma solução de código aberto, proporcionando mais economia que soluções pagas. Além disso, o *PostgreSQL* é capaz de realizar a grande maioria das transações *SQL* e controle simultâneo. *PostgreSQL* também permite adicionar extensões de tipo de dados, funções, operadores e linguagens processuais e oferece recursos como consultas complexas, integridade transacional e gatilhos (BOEHM; ABTS; CHULANI, 2000; BRIAND; WIECZOREK, 2002).

2.5 Kanban

Nos anos 1990 o desenvolvimento de *software* encontrava muitas barreiras para produção de *softwares* de forma mais rápida e eficiente. Em 2001 foi criado o manifesto ágil que deu origem aos métodos ágeis de desenvolvimento de *software*. Os métodos ágeis defendem entregas constantes e resultados rápidos sempre visando a qualidade do produto e a satisfação do cliente. Os métodos ágeis também possuem documentação moderada, iterações curtas e alterações imediatas (PONTES et al., 2014; OUHBI; POMBO, 2020). A Figura 1 apresenta os doze princípios dos métodos ágeis. Esses princípios foram adotados por várias metodologias ágeis que ganharam muita popularidade como o *Scrum* (STOPA; RACHID, 2019), *Extreme Programming (XP)* (LINDSTROM; JEFFRIES, 2003) e *Kanban* (BOEG, 2010).

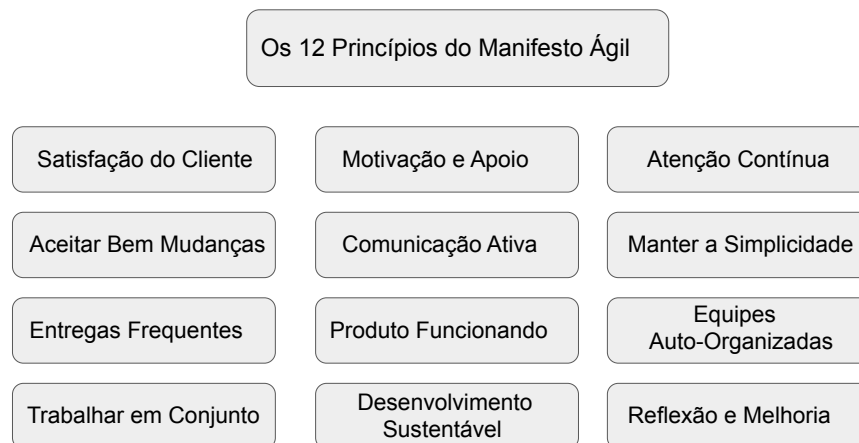


Figura 1 – Os 12 princípios do manifesto ágil.

O *Kanban* é um dos métodos ágeis mais populares e pode ser usado até para desenvolvimento de outros produtos, além de *softwares*. A origem do *Kanban* foi na década de 1950 na empresa *Toyota* em um sistema de produção *JIT (Just-In-Time)*, que defende fazer apenas o necessário, quando é necessário e na quantidade necessária. Em 2004 a ideia do *Kanban* foi aplicada para o ramo de desenvolvimento de *software* em uma equipe

mal engajada de TI (Tecnologia da Informação) na empresa *Microsoft* (AHMAD; MARK-KULA; OIVO, 2013). Com o *Kanban* as equipes podem visualizar o fluxo de trabalho, limitar o trabalho em progresso (*WIP*) em cada estágio do fluxo de trabalho e reduzir o tempo de produção. A Figura 2 apresenta um exemplo do quadro *Kanban*. Esse quadro fornece visibilidade para o processo de software, dando *feedback* do que cada desenvolvedor está fazendo, além de comunicar claramente prioridades, destaques e gargalos. O foco do *Kanban* é reduzir o tempo de produção do *software* focando no fluxo e na ausência da obrigatoriedade de iterações (KIROVSKA; KOCESKI, 2015; CORONA; PANI et al., 2013).

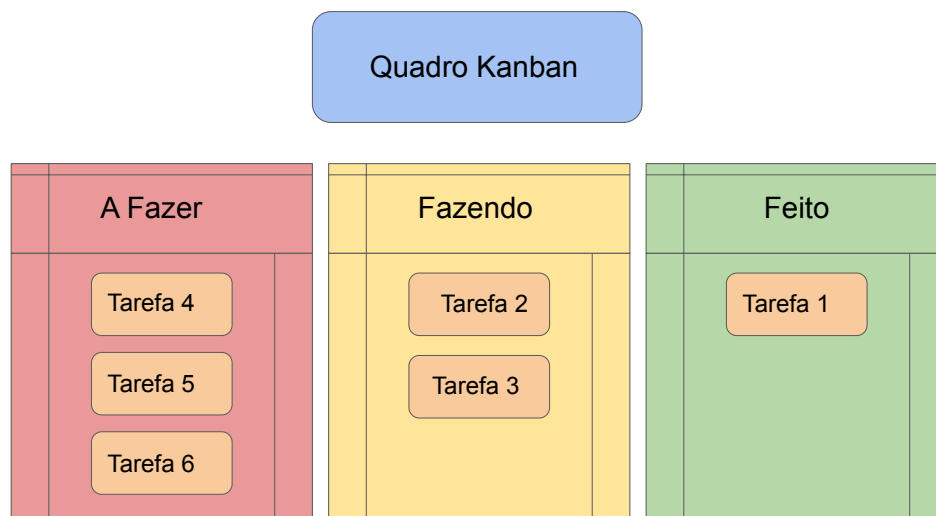


Figura 2 – Exemplo de quadro Kanban.

2.6 Django

O *Django* é um *framework* para desenvolvimento *web* utilizado em conjunto com a linguagem de programação *python* para criar *websites* de forma mais rápida e eficiente. Um *framework* é uma ferramenta para desenvolvimento de *software* que aproveita certos padrões em vários projetos de software para entregar determinadas funcionalidades. O *Django* foi criado em 2003 em uma assessoria de imprensa do Kansas. Em 2005 a agência decidiu publicar o código fonte do *Django* e fundaram a *Django Software Foundation* em 2008 para dar mais suporte ao *framework*. A primeira versão do *software* foi lançada em 2008. O *Django* utiliza o padrão *Model Template View (MTV)* para construir uma arquitetura completa (DAUZON; BENDORAITIS; RAVINDRAN, 2016).

Django é utilizado para organizar a estrutura da aplicação, além de automatizar e agilizar diversas fases do processo de criação de *websites*. A estrutura do *Django* foi construída em conectividade de banco de dados para facilitar o armazenamento de dados e o acesso aos dados. O *Django* permite integração com vários bancos de dados diferentes,

fazendo toda comunicação e manipulação dos dados através de código *Python*. O *Django* permite que o código *python* seja incorporado em *html* da página. Além disso, o *Django* fornece um site com um sistema de administração integrado, que pode ser usada para gerenciar usuários e tabelas do banco de dados ([VAMSI et al., 2021](#)).

3 Trabalhos Relacionados

Este capítulo apresenta os trabalhos relacionados que possuem propostas alinhadas com a melhoria de qualidade de rebanhos bovinos.

[Abdalla \(2003\)](#) propõem melhorar o desempenho reprodutivo de vacas búfalas. Abdalla et al. usaram sessenta e quatro vacas búfalas multíparas egípcias criadas sob dois sistemas de manejo diferentes em duas fazendas que foram divididas em dois grupos e continuamente expostos a um touro fértil seguindo estratégias para cada grupo. Os resultados mostraram que os intervalos pós-parto para a primeira ovulação, primeiro estro, concepção e próximo parto no primeiro grupo foram bem mais curtos do que no segundo grupo.

[Brumatti et al. \(2011\)](#) propõem um modelo bio-econômico para determinação de ponderadores econômicos utilizados em índices de seleção em programas de melhoramento genético de gado. Os resultados de Brumatti et al. mostram que a soma das características reprodutivas foi mais importantes do que as características de crescimento, mostrando as necessidades de se reformular os índices de seleção.

O trabalho de [Chandra, Sejian e Sharma \(2015\)](#) apresentam algumas estratégias para melhorar a reprodução pecuária através de mudanças climáticas. Chandra et al. defendem que é possível melhorar a capacidade de um animal lidar com o estresse ambiental através do manejo estratégico da reprodução pela manipulação da foliculogênese, alterações hormonais, reprodução seletiva e aplicação de técnicas de transferência de embriões.

[Fontoura et al. \(2019\)](#) apresentam uma solução baseada em algoritmos genéticos que analisa o conjunto principal das variáveis envolvidas para recomendar acasalamentos ao produtor. Os resultados de Fontoura et al. apontam um desempenho satisfatório dos algoritmos expondo a inclinação pela escolha dos animais de maior valor das diferenças esperadas de progênie com base nas ponderações determinadas pelo produtor.

O trabalho de [Keshavarzi et al. \(2020\)](#) apresentam uma solução para prever a perda de gravidez em rebanhos leiteiros iranianos a fim de aumentar a lucratividade da fazenda. Keshavarzi et al. usam um software de manutenção de registros na fazenda que analisa o histórico das vacas e informações genéticas de touros disponíveis em 6 grandes fazendas leiteiras comerciais com vacas paridas entre 2005 e 2014. O resultado esperado era acima de 90%, entretanto os modelos de predição utilizados variaram de 32,3% a 69,2%.

O trabalho de [Lobo \(2021\)](#) apresentam os principais fatores para melhoria de bovinos de corte no Brasil. Lobo et al. aborda questões e funcionamento do processo de seleção de bovinos.

[Motta, Ferreira e Yokoo \(2021\)](#) produzem uma estratégia ótima de acasalamentos entre animais selecionados para reprodução através do algoritmo *BrangusSelection*. O algoritmo de Da et al. depende de um índice de seleção, obtido a partir dos objetivos e critérios de

seleção dos animais. Os testes com o *BrangusSelection* mostram que o tempo de execução é polinomial de acordo com a quantidade de animais recebidos como entrada, o que indica que o problema pode ser tratado computacionalmente.

[Andrietta \(2022\)](#) propõem uma otimização de acasalamentos através de uma simulação de uma população de bovinos leiteiros e da exploração de diferentes abordagens de extração de atributos de informações genotípicas de indivíduos do rebanho. Andrietta et al. usam dois algoritmos de *Machine Learning* em onze cenários propostos referentes ao coeficiente de endogamia, valor genético, além da proposta de um método de acasalamento. Os resultados mostram que a predição do coeficiente de endogamia e valor genético potencial são uma estratégia de direcionamento de acasalamentos para otimizar o ganho genético e controlar a endogamia.

A Tabela 1 resume um comparativo dos trabalhos relacionados sob quatro aspectos: Ano, Objetivo, Ciência de Dados e Matriz de Parentesco. O primeiro critério focou-se no **Ano** de publicação dos trabalhos relacionados. Os trabalhos foram publicados entre 2003 a 2022, principalmente no período de 2019 a 2021 mostrando que o tema tem ganhando cada vez mais relevância. O segundo critério aborda o objetivo dos trabalhos relacionados. Todos os trabalhos relacionados tem um objetivo alinhado a melhorar a reprodução do rebanho.

O terceiro critério aborda a utilização da Ciência de Dados nas propostas dos trabalhos relacionados. A Ciência de Dados pode ser uma ferramenta poderosa para melhoria genética de rebanhos, treinando algoritmos para sugerir diversas possibilidades de melhoria genética. Apenas cinco trabalhos relacionados utilizaram a ciência de dados em suas propostas.

O quarto critério foca-se na utilização da Matriz de Parentesco nas propostas dos trabalhos relacionados. A Matriz de Parentesco pode ser usada para guardar os dados da árvore genealógica dos gados. [Fontoura et al. \(2019\)](#), [Fontoura \(2019\)](#) abordam a matriz de parentesco para sugerir os melhores parceiros para cruzamentos, evitando a perda de qualidade do rebanho. Apenas quatro trabalhos relacionados utilizaram a matriz de parentesco em suas propostas.

Tabela 1 – Comparação de trabalhos a partir dos ano, objetivo, ciência de dados e matriz de parentesco.

Trabalho	Objetivo	Ciência de Dados	Matriz de Parentesco
Abdalla (2003)	Melhorar o desempenho reprodutivo de vacas búfalas	Não	Não
Brumatti et al. (2011)	Determinar ponderadores econômicos utilizados em índices de seleção em programas de melhoramento genético de bovinos de corte	Não	Não
Chandra, Sejian e Sharma (2015)	Estratégias para melhorar a reprodução pecuária através de mudanças climáticas	Não	Não
Fontoura et al. (2019)	Otimizar a seleção de parceiros em programas de melhoramento genético de bovinos de corte	Sim	Sim
Keshavarzi et al. (2020)	Prever a perda de gravidez em rebanhos leiteiros iranianos	Sim	Não
Lobo (2021)	Apresentar os principais fatores envolvidos no processo de melhoramento de bovinos de corte no Brasil	Não	Não
Motta, Ferreira e Yokoo (2021)	Produzir uma estratégia ótima de acasalamentos entre animais selecionados para reprodução	Sim	Sim
Andrietta (2022)	Otimizar acasalamentos	Sim	Sim
Trabalho Proposto	Auxiliar a tomada de decisões de cruzamentos para melhorar a qualidade genética do rebanho	Sim	Sim

4 Aplicação Desenvolvida

Este capítulo aborda aspectos relevantes ao desenvolvimento das funcionalidades do sistema que foram desenvolvidos no decorrer deste trabalho, com ênfase na ciência de dados relacionados à matriz de parentesco dos animais.

4.1 Metodologia

A metodologia utilizada no desenvolvimento da aplicação desta monografia consistiu em: levantamento bibliográfico; levantamento e análise de requisitos; desenvolvimento dos algoritmos e integração com o sistema; testes de conformidade; validação; e implantação. A Figura 3 apresenta a metodologia descrita acima por meio de um modelo baseado em UML (*Unified Modeling Language*)

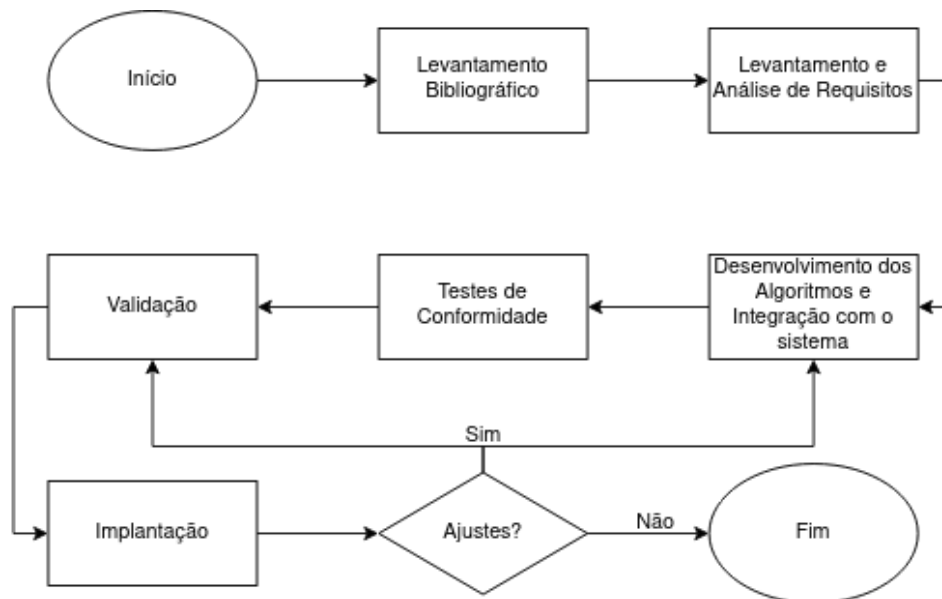


Figura 3 – Fluxograma de desenvolvimento.

Levantamento Bibliográfico: Essa é a primeira fase do projeto, onde é feito um levantamento inicial da literatura a fim de identificar o problema corretamente e verificar se ele já não foi solucionado. O levantamento também é responsável por apresentar métricas complementares que possam ser úteis para o sistema

Levantamento e Análise de Requisitos: Nessa fase busca-se compreender o sistema que será desenvolvido e definir o escopo da aplicação, bem como ela irá funcionar, de onde virão os dados, o que será feito com eles e para onde irão. Nessa fase é gerada a tabela de requisitos funcionais e não funcionais, que serão vistas nas próximas seções.

Desenvolvimento dos Algoritmos e Integração com o Sistema: Nessa fase é dado início ao desenvolvimento do sistema. A metodologia de desenvolvimento adotada foi baseada na utilizada por [Martins \(1997\)](#). O foco principal desta monografia é a geração dos algoritmos de matriz de parentesco para serem integrados em um sistema existente, então nessa fase também há um esforço para integrá-los ao sistema principal.

Testes de Conformidade: Nessa fase são realizados alguns testes que visam identificar se os dados que estão sendo gerados pelos algoritmos estão em conformidade com o que é esperado.

Validação: Nessa fase o sistema como um todo é verificado a fim de verificar se está se comportando como esperado e se há o cumprimento de todos os requisitos estabelecidos nas etapas iniciais do projeto.

Implantação: Nessa fase o sistema será implantando e disponibilizado para acesso aos usuários. Nessa fase podem haver erros que não foram percebidos em outras etapas, então, caso sejam necessários ajustes, deve regressar algumas etapas para correção e/ou validação até que o sistema esteja verdadeiramente pronto.

4.2 Requisitos Funcionais

A Tabela 2 apresenta os requisitos funcionais do sistema. Os requisitos funcionais de um sistema são as especificações detalhadas das funções, tarefas e operações que o sistema deve ser capaz de realizar para atender às necessidades e expectativas dos usuários. Em outras palavras, os requisitos funcionais definem o comportamento esperado do sistema e as tarefas que ele deve ser capaz de executar. Os requisitos funcionais são identificados durante a fase de análise de requisitos do ciclo de vida de desenvolvimento de software. Eles descrevem o que o sistema deve fazer, em vez de como ele deve ser implementado. A partir desses requisitos, os desenvolvedores projetam o sistema e criam um plano de desenvolvimento para implementar cada um dos requisitos. O foco principal deste projeto está concentrado nos requisitos que vão do 09 ao 18, em outras palavras, os requisitos que envolvem a geração da matriz de parentesco.

Tabela 2 – Requisitos funcionais.

Identificador	Descrição	Dependência
RF01: Cadastro de Usuário	Usuários podem ser cadastrados como criadores, técnicos, superintendente e administradores.	-
RF02: <i>Login</i>	O sistema permitirá o acesso à plataforma aos usuários, autenticados com e-mail e senha.	RF01

RF03: Acesso do usuário	O criador poderá realizar uma comunicação de cobertura e comunicação de nascimento, além de acessar os dados de todos os animais.	RF01, RF02
RF04: Acesso do usuário	O técnico poderá visualizar, registrar novos animais e aprovar cobertura de nascimento.	RF01, RF02
RF05: Acesso do usuário	O Superintendente poderá criar, listar, atualizar e deletar usuários.	RF01, RF02
RF06: Acesso do usuário	O Superintendente poderá aprovar comunicação de nascimento, de coberturas e aprovar registro de animais realizado pelo técnico.	RF01, RF02
RF07: Acesso do usuário	O Superintendente poderá aprovar cadastro de criadores.	RF01, RF02
RF08: Acesso do usuário	Os usuários administradores poderão manipular todas as tabelas do banco de dados através da visualização administrativa.	-
RF09: Grau de endogamia	Calcular a matriz de parentesco dos animais com registro no sistema. (MARTINS, 1997)	-
RF10: Grau de endogamia	Retornar o parentesco entre dois quaisquer animais i e j de acordo com Matriz de Parentesco. Ver Equação 4.1	RF09
RF11: Grau de endogamia	Retornar o Parentesco médio da raça. Ver Equação 4.2	RF09
RF12: Grau de endogamia	Retornar o coeficiente de endogamia de um animal específico. Ver Equação 4.3	RF09
RF13: Grau de endogamia	Retornar o coeficiente de endogamia médio da raça. Ver Equação 4.4	RF09
RF14: Grau de endogamia	Retornar o coeficiente de endogamia médio por rebanho. Ver Equação 4.5	RF09
RF15: Grau de endogamia	Retornar a endogamia por ano de nascimento dos animais. Ver Equação 4.6	RF09
RF16: Grau de endogamia	Calcular o intervalo de gerações geral. Ver Equação 4.7	RF09
RF17: Grau de endogamia	Calcular o intervalo de gerações dos animais reprodutores. Ver Equação 4.8	RF09
RF18: Grau de endogamia	Calcular o intervalo de gerações dos animais matrizes. Ver Equação 4.9	RF09
RF19: Arvore genealógica	Criar arvore genealógica para um animal específico.	RF03, RF04, RF05

RF20:	Dados dos animais	Exportar arquivo de dados: Número do Animal, Número do Pai do animal, Número da Mãe do animal, Data de Nascimento, Sexo e Rebanho.	RF03, RF04, RF05
-------	-------------------	--	------------------

4.3 Requisitos Não Funcionais

Os requisitos não funcionais de um sistema indicam especificações que estão relacionadas à qualidade e comportamento do sistema. A Tabela 3 apresenta os requisitos não funcionais do sistema.

Tabela 3 – Requisitos Não Funcionais.

Identificador	Descrição	Categoria
RNF01	É necessário implementar um sistema de autenticação e autorização que restrinja a manipulação dos dados somente para usuários logados e autorizados.	Segurança
RNF02	É fundamental que o sistema mantenha a integridade e segurança dos dados no banco de dados, impedindo qualquer tipo de manipulação indevida.	Segurança
RNF03	A interface do sistema deve ser simples e intuitiva para facilitar a sua utilização.	Usabilidade
RNF04	O sistema precisa garantir alta disponibilidade, permanecendo disponível o tempo todo.	Eficiência
RNF05	O sistema deve ter um layout responsivo para se adaptar a diferentes tamanhos de tela, garantindo a sua portabilidade.	Portabilidade

4.4 Modelagem do Banco de Dados

A Figura 4 apresenta o esquema do banco de dados da aplicação como um todo. Por questões de simplificação, as tabelas contam apenas com suas chaves primárias e estrangeiras, com exceção da tabela `adminApp_matrix` que representa os dados relativos à matriz de parentesco. As tabelas `superintendente`, `criador` e `tecnico` representam os tipos de usuários do sistema (tabela `usuario`), cada um com suas permissões. As demais tabelas, com exceção da tabela de matriz de parentesco, representam dados relativos ao registro dos animais. A tabela `comunic_cobert` representa um cruzamento entre animais, enquanto `comunic_nasci` representa o nascimento de um animal. A tabela `fazenda` representa os dados de uma fazenda criadora de gado, em outras palavras, um rebanho. Por fim, a tabela `animais_animal` representa os dados de um animal específico, incluindo a informação as

chaves estrangeiras para o seu rebanho, pai e mãe, bem como a data de nascimento do animal, que são dados relevantes para os cálculos do trabalho.

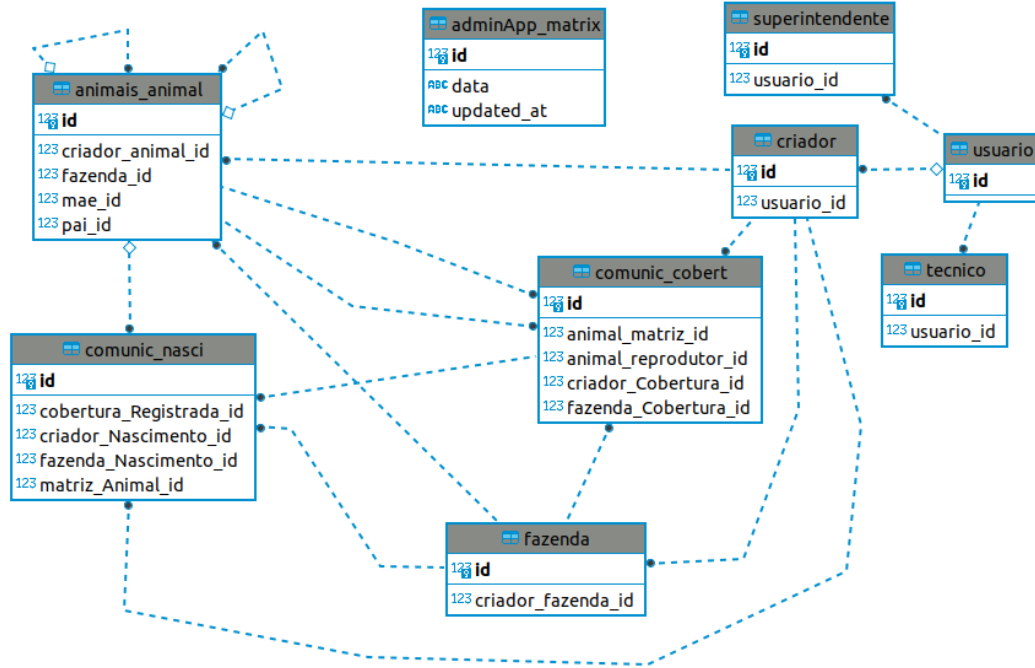


Figura 4 – Esquema simplificado de tabelas do banco de dados.

4.5 Métricas Calculadas

A Equação 4.1 representa o grau de parentesco (GrP) entre 2 animais quaisquer no conjunto de animais registrados no sistema, para a obtenção desse valor, basta olhar para a matriz de parentesco com base no índice dos animais, por exemplo, o parentesco entre o animal de índice 1 e o animal de índice 2 é o elemento da linha 1 e coluna 2 da matriz ou o elemento da linha 2 coluna 1, tendo em vista que a matriz de parentesco é sempre quadrada e simétrica em relação à diagonal principal.

$$GrP = a_{ij}. \quad (4.1)$$

A Equação 4.2 representa o grau de parentesco médio (GrPm) da raça, sendo dado como a média dos valores da matriz de parentesco que não estão contidos na diagonal principal, totalizando $n * (n - 1)$ elementos e sendo cada elemento representado por a_{ij} quando i é diferente de j

$$GrPm = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n a_{ij}. \quad (4.2)$$

A Equação 4.3 representa o coeficiente de endogamia (F) de um animal em específico, sendo esse calculado como a subtração do elemento da diagonal principal da matriz de parentesco correspondente ao índice do animal subtraído do valor 1.

$$F = a_{ii} - 1. \quad (4.3)$$

A Equação 4.4 representa o coeficiente de endogamia médio (Fm), dado como a média do valor F de todos os animais do sistema.

$$Fm = \frac{1}{n} \sum_{i=1}^n (a_{ii} - 1). \quad (4.4)$$

A Equação 4.5 representa o coeficiente de endogamia médio de um rebanho r (Fm_r) dado como a média do valor F dos animais pertencentes ao rebanho r , sendo b a submatriz com apenas os animais do rebanho.

$$Fm_r = \frac{1}{k} \sum_{i=1}^k (b_{ii} - 1). \quad (4.5)$$

A Equação 4.6 representa o coeficiente de endogamia médio de um ano a (Fm_a) dado como a média do valor F dos animais nascidos no ano a , sendo c a submatriz com apenas os animais nascidos naquele ano.

$$Fm_a = \frac{1}{l} \sum_{i=1}^l (c_{ii} - 1). \quad (4.6)$$

A Equação 4.7 representa o valor do intervalo de gerações (IG) dos animais e é dado como a média dos valores de idade dos animais no momento do nascimento de seus filhotes em anos, sendo d o conjunto de idades dos animais reprodutores e e o conjunto de idades dos animais matrizes.

$$IG = \frac{1}{p + m} \left(\sum_{i=1}^p d_i + \sum_{i=1}^m e_i \right). \quad (4.7)$$

A Equação 4.8 representa o valor do intervalo de gerações de animais reprodutores IG_m e é dado como a média dos valores de idade dos animais do sexo masculino no momento do nascimento de seus filhotes em anos.

$$IG_m = \frac{1}{p} \left(\sum_{i=1}^p d_i \right). \quad (4.8)$$

A Equação 4.9 representa o valor do intervalo de gerações de animais matrizes (IG_f) e é dado como a média dos valores de idade dos animais do sexo feminino no momento do nascimento de seus filhotes em anos.

$$IG_f = \frac{1}{m} \left(\sum_{i=1}^m e_i \right). \quad (4.9)$$

5 Resultados e Discussões

Este capítulo aborda um exemplo de conjunto de dados pensado para validar a realização do cálculo das métricas. Para tanto, serão apresentadas 2 seções, a primeira contendo os dados iniciais relativos à ancestralidade, a matriz de parentesco e os dados extraídos destes para a validação dos cálculos, enquanto a segunda servirá para mostragem dos dados da forma que são representados no sistema.

5.1 Dados Preliminares

Na Tabela 4 temos um conjunto de dados simplificado contendo informações do índice representativo de cada animal, bem como o índice de seu pai e mãe, com o valor 0 indicando a ausência de informações sobre, além da data de nascimento, que é necessária para o cálculo do intervalo de gerações da raça (4.7).

Tabela 4 – Tabela de *Pedigree* com Data de Nascimento e Rebanho.

Animal	Pai	Mãe	Data de Nascimento	Rebanho
1	0	0	01/01/2016	Fazenda 1
2	0	0	01/01/2017	Fazenda 1
3	0	0	01/01/2018	Fazenda 1
4	1	2	01/01/2019	Fazenda 1
5	1	2	01/01/2020	Fazenda 2
6	1	2	01/01/2021	Fazenda 2
7	4	6	01/01/2022	Fazenda 2
8	3	5	01/01/2023	Fazenda 2

Na matriz 5.1 temos os dados da matriz de parentesco de toda a raça, conforme representado no *pedigree* 4, calculado conforme o método descrito por (MARTINS, 1997). Pode-se observar uma matriz quadrada de ordem 8 e simétrica em relação à sua diagonal principal, que representa a endogamia de cada animal, enquanto os elementos fora da diagonal principal representam o grau de consanguinidade de cada par de animais. Através desses dados será possível calcular todas as demais métricas apresentadas neste trabalho.

$$\begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.500 & 0.500 & 0.500 & 0.500 & 0.250 \\
0.000 & 1.000 & 0.000 & 0.500 & 0.500 & 0.500 & 0.500 & 0.250 \\
0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.500 \\
0.500 & 0.500 & 0.000 & 1.000 & 0.500 & 0.500 & 0.750 & 0.250 \\
0.500 & 0.500 & 0.000 & 0.500 & 1.000 & 0.500 & 0.500 & 0.500 \\
0.500 & 0.500 & 0.000 & 0.500 & 0.500 & 1.000 & 0.750 & 0.250 \\
0.500 & 0.500 & 0.000 & 0.750 & 0.500 & 0.750 & 1.250 & 0.250 \\
0.250 & 0.250 & 0.500 & 0.250 & 0.500 & 0.250 & 0.250 & 1.000
\end{bmatrix} \quad (5.1)$$

A matriz 5.2 representa o subconjunto da matriz de parentesco completa que corresponde aos animais da Fazenda 1 levando em conta a ancestralidade de todos os animais do *pedigree*. Os dados importantes são os elementos da diagonal principal, por meio do qual foi calculado o Fm_1 .

$$\begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.500 \\
0.000 & 1.000 & 0.000 & 0.500 \\
0.000 & 0.000 & 1.000 & 0.000 \\
0.500 & 0.500 & 0.000 & 1.000
\end{bmatrix} \quad (5.2)$$

A matriz 5.3 representa o subconjunto da matriz de parentesco completa que corresponde aos animais da Fazenda 2 levando em conta a ancestralidade de todos os animais do *pedigree*. Os dados importantes são os elementos da diagonal principal, por meio dos quais foi calculado o Fm_2 .

$$\begin{bmatrix}
1.000 & 0.500 & 0.500 & 0.500 \\
0.500 & 1.000 & 0.750 & 0.250 \\
0.500 & 0.750 & 1.250 & 0.250 \\
0.500 & 0.250 & 0.250 & 1.000
\end{bmatrix} \quad (5.3)$$

A lista 5.4 representa o conjunto de dados correspondente à idade dos animais reprodutores no momento do nascimento de seus filhos, extraído como a diferença das datas de nascimento do animal nascido e de seu pai em anos. Os dados estão organizados de acordo com a sequência do *pedigree*, sendo o primeiro elemento a idade do animal reprodutor no nascimento do animal de índice 4 e o último, do animal de índice 8.

$$\begin{bmatrix} 3 & 4 & 5 & 3 & 5 \end{bmatrix} \quad (5.4)$$

A lista 5.5 representa o conjunto de dados correspondente à idade dos animais matrizes no momento do nascimento de seus filhos, extraído como a diferença das datas de nascimento do animal nascido e de sua mãe em anos. Os dados estão organizados de acordo com a sequência do *pedigree*, sendo o primeiro elemento a idade do animal reprodutor no nascimento do animal de índice 4 e o último, do animal de índice 8.

Figura 5 – Tela de mostragem do parentesco entre 2 animais.

Outro dado relevante consiste no subconjunto da matriz de parentesco representativo dos animais nascidos a cada ano, ou simplesmente o conjunto com os elementos da diagonal principal, no entanto, o conjunto de dados representa apenas um animal nascido por ano, levando a cada subconjunto ser representado pelo elemento da diagonal principal correspondente a seu índice. Por tal motivo, o subconjunto de cada ano está omitido de representação figurativa.

$$\begin{bmatrix} 2 & 3 & 4 & 1 & 3 \end{bmatrix} \quad (5.5)$$

5.2 Representação dos Dados

A Figura 5 representa a mostragem dos dados de parentesco entre 2 quaisquer animais cadastrados no sistema. Para tanto, tem-se 2 seletores que listam os animais nos quais é possível escolher 2 deles, que o parentesco será mostrado no campo logo a baixo. No exemplo, foram selecionados os animais de índice 1 e 2, correspondendo ao elemento a_{67} ou ao elemento a_{76} da matriz de parentesco, correspondendo ao valor 0.75.

Na Figura 6 temos a representação do valor de parentesco médio da raça bovina, ou seja, de todos os animais cadastrados, que é dada como a média de todos os valores fora da diagonal principal da matriz de parentesco, cuja soma resulta no valor 19.5 que dividido por 56 resulta no valor de GrPm em aproximadamente 0.348.

Na Figura 7 temos a representação do valor do coeficiente médio de endogamia da raça bovina, dado como a média aritmética dos coeficientes de endogamia (4.3), de todos os animais cadastrados, ou seja, a média do valor calculado do elemento a_{ii} subtraído

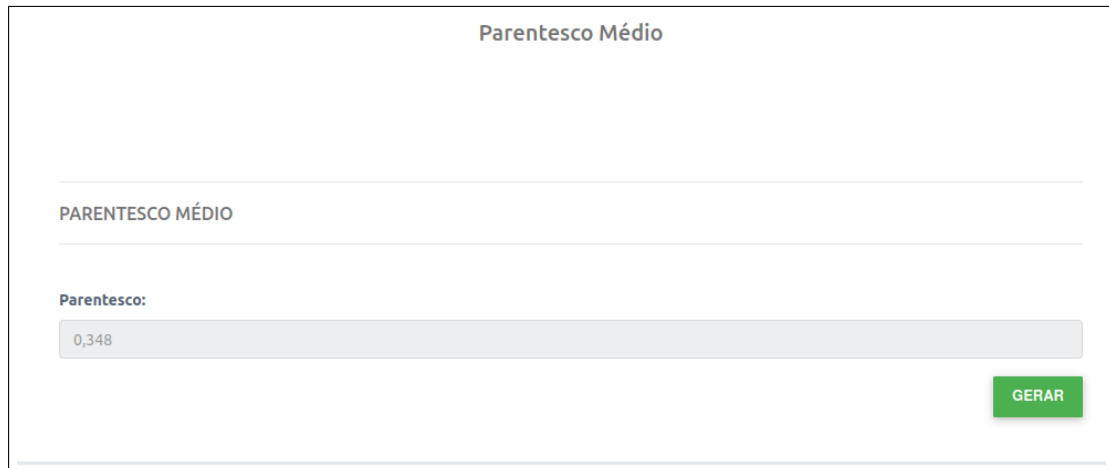


Figura 6 – Tela de mostragem do parentesco médio da raça.

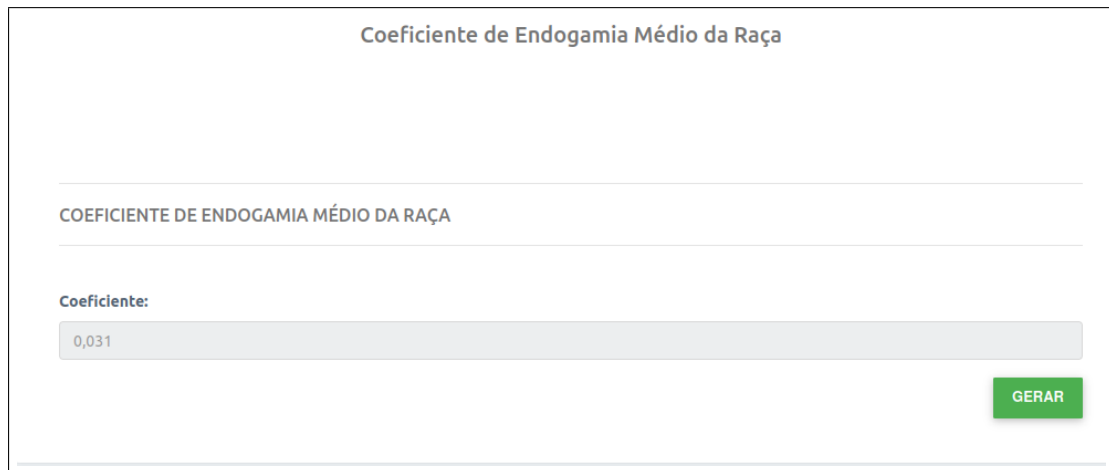


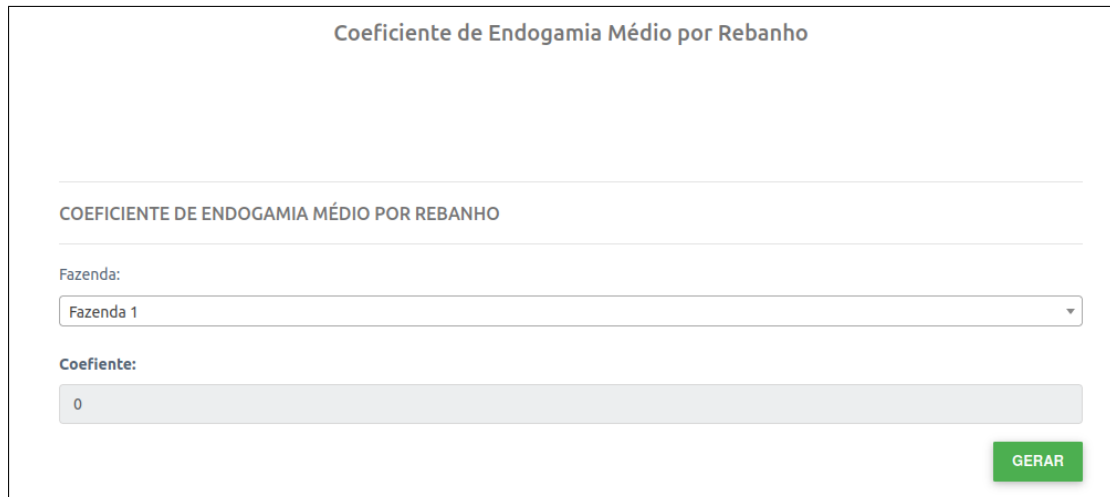
Figura 7 – Tela de mostragem do coeficiente de endogamia médio da raça.

de 1, cuja soma resulta no valor 0.25 que dividido por 8 resulta no valor de F_m em aproximadamente 0.031.

Na Figura 8 temos a representação do valor do coeficiente médio de endogamia dos animais da Fazenda 1, dado como a média aritmética dos valores F dos animais da Fazenda 1, ou seja, a média do valor calculado do elemento b_{ii} subtraído de 1 da matriz de parentesco da Fazenda 1 5.2, cuja soma resulta no valor 0 que dividido por 4 resulta no valor de F_{m1} em 0.

Na Figura 9 temos a representação do valor do coeficiente médio de endogamia dos animais da Fazenda 2, dado como a média aritmética dos valores F dos animais da Fazenda 2, ou seja, a média do valor calculado do elemento c_{ii} subtraído de 1 da matriz de parentesco da Fazenda 2 5.3, cuja soma resulta no valor 0.25 que dividido por 4 resulta no valor de F_{m2} em 0.062 aproximadamente.

Na Figura 10 temos um gráfico representativo do valor F_{m_a} a cada ano, calculado como o valor médio do coeficiente de endogamia dos animais nascidos em cada ano específico, e na situação proposta temos um animal por ano, formulando o valor de F a cada ano,



Coeficiente de Endogamia Médio por Rebanho

COEFICIENTE DE ENDOGAMIA MÉDIO POR REBANHO

Fazenda:

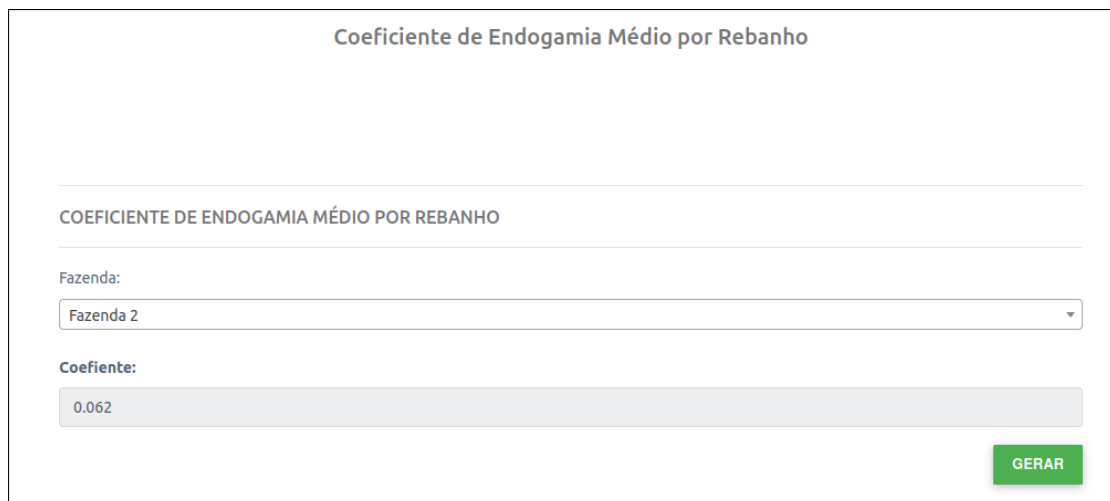
Fazenda 1

Coeficiente:

0

GERAR

Figura 8 – Tela de mostragem do coeficiente de endogamia médio da Fazenda 1.



Coeficiente de Endogamia Médio por Rebanho

COEFICIENTE DE ENDOGAMIA MÉDIO POR REBANHO

Fazenda:

Fazenda 2

Coeficiente:

0.062

GERAR

Figura 9 – Tela de mostragem do coeficiente de endogamia médio da Fazenda 2.

sendo que apenas no ano de 2022 temos um valor diferente de zero, tendo em vista que o animal 7 é o único animal endogâmico, ou seja, a_{77} maior que 1.

Na Figura 11 temos um gráfico representativo do valor Fm_a a cada ano dos animais da Fazenda 1, calculado como o valor médio do coeficiente de endogamia dos animais nascidos em cada ano específico e que estão no rebanho da Fazenda 1, na situação temos os valores correspondentes ao valor de F para primeiros 4 animais do *pedigree* nos anos de 2016 a 2019, que são os valores zero em todos eles.

Na Figura 12 temos um gráfico representativo do valor Fm_a a cada ano dos animais da Fazenda 2, calculado como o valor médio do coeficiente de endogamia dos animais nascidos em cada ano específico e que estão no rebanho da Fazenda 2, na situação temos os valores correspondentes ao valor de F para últimos 4 animais do *pedigree* nos anos de 2020 a 2023, que constitui apenas o valor 0.25 no ano de 2022, correspondendo ao animal 7.

Na Figura 13 temos os valores de IG_m dado como a média aritmética da lista 4.8,

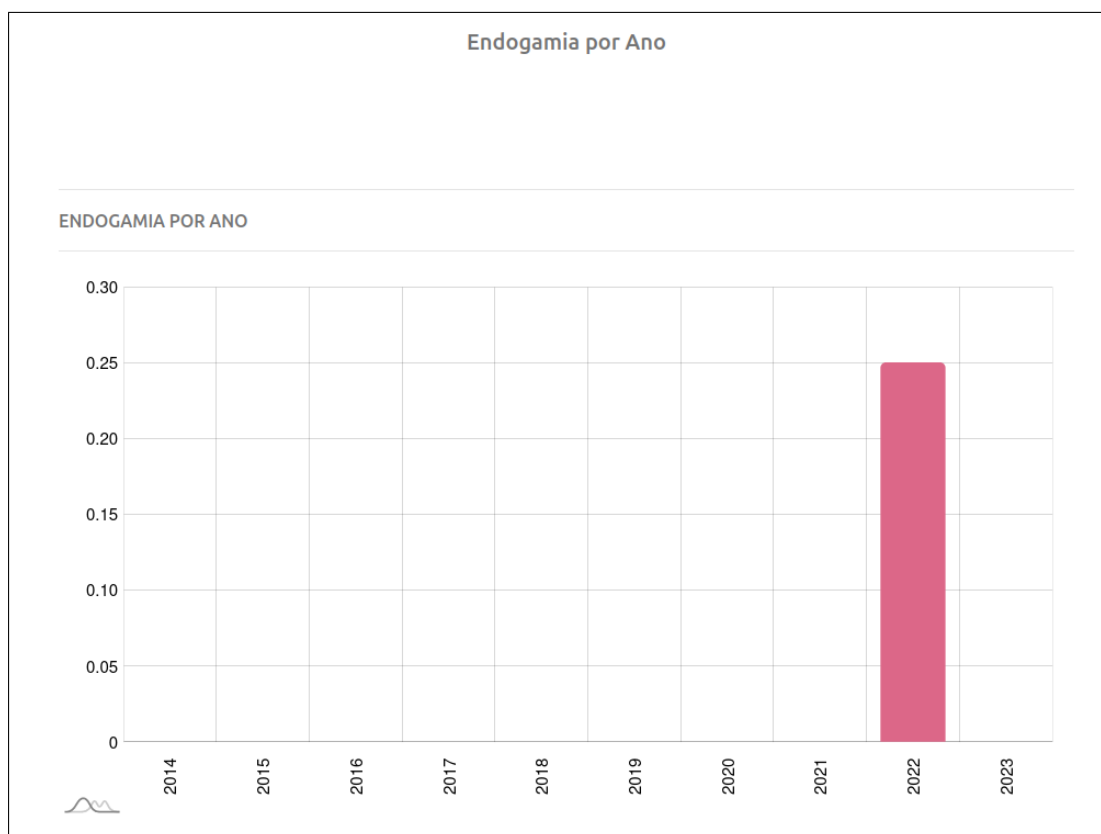


Figura 10 – Tela de mostragem do coeficiente de endogamia agrupado por ano de nascimento.

cuja soma corresponde a 20, que dividido por 5 representa o valor 4 como resultado. Em animais matrizes (IG_f), dado como a média aritmética da lista 4.9, cuja soma corresponde a 13, que dividido por 5 representa o valor 2.6 como resultado. Por fim, temos o valor de IG , dado como a média aritmética da combinação das duas listas já citadas, cuja soma corresponde a 33, que dividido por 10 representa o valor 3.3 como resultado.

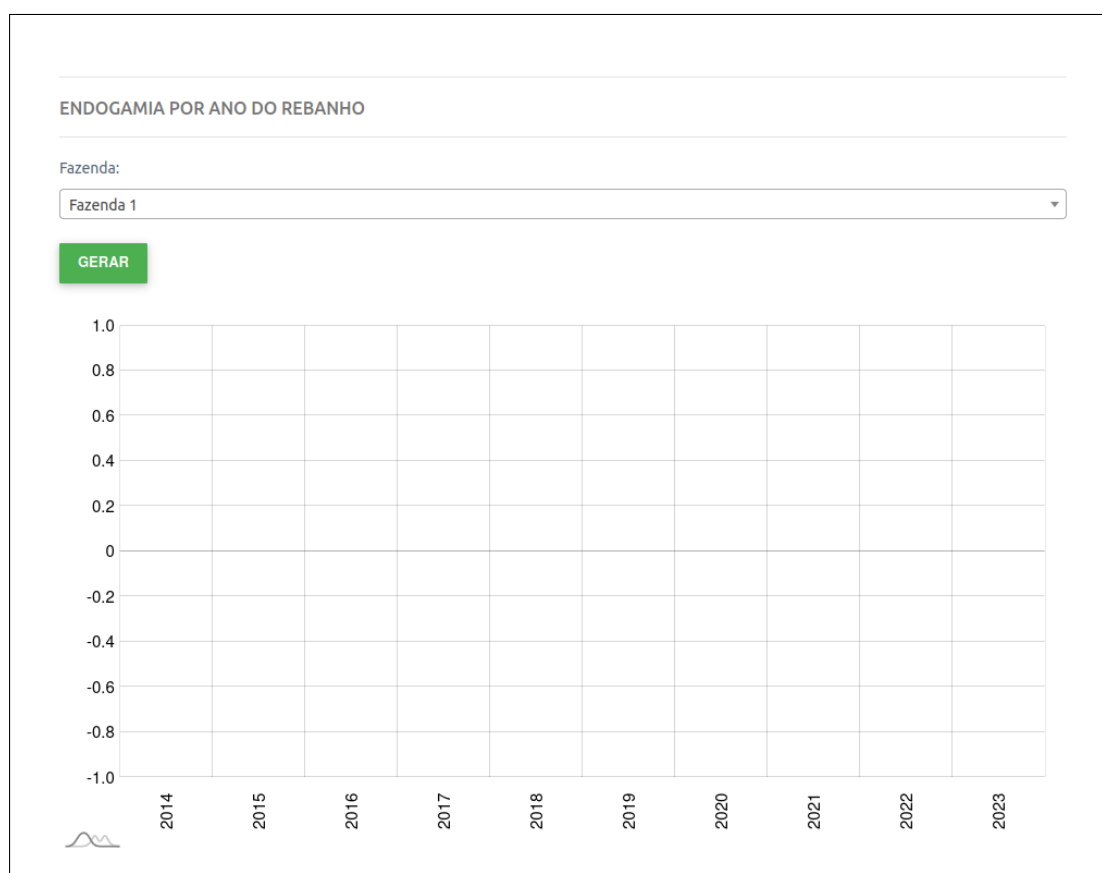


Figura 11 – Tela de mostragem do coeficiente de endogamia agrupado por ano de nascimento da Fazenda 1.

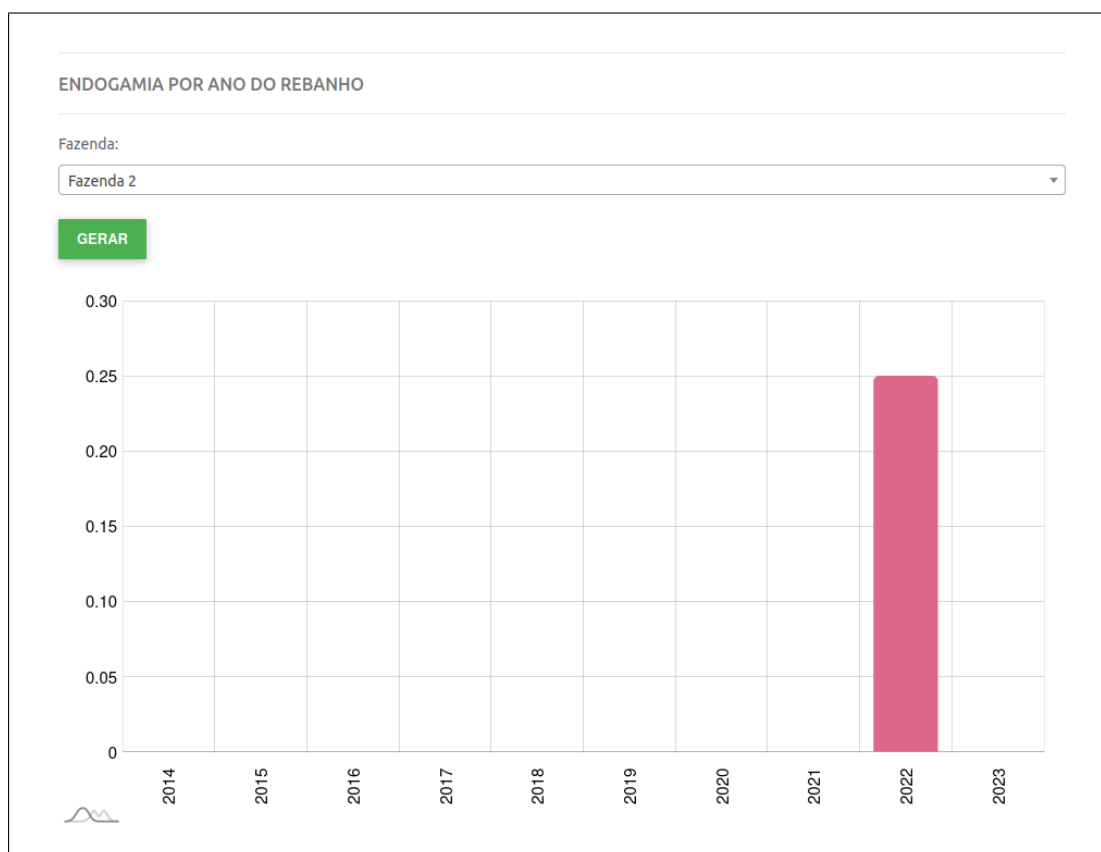


Figura 12 – Tela de mostragem do coeficiente de endogamia agrupado por ano de nascimento da Fazenda 2.

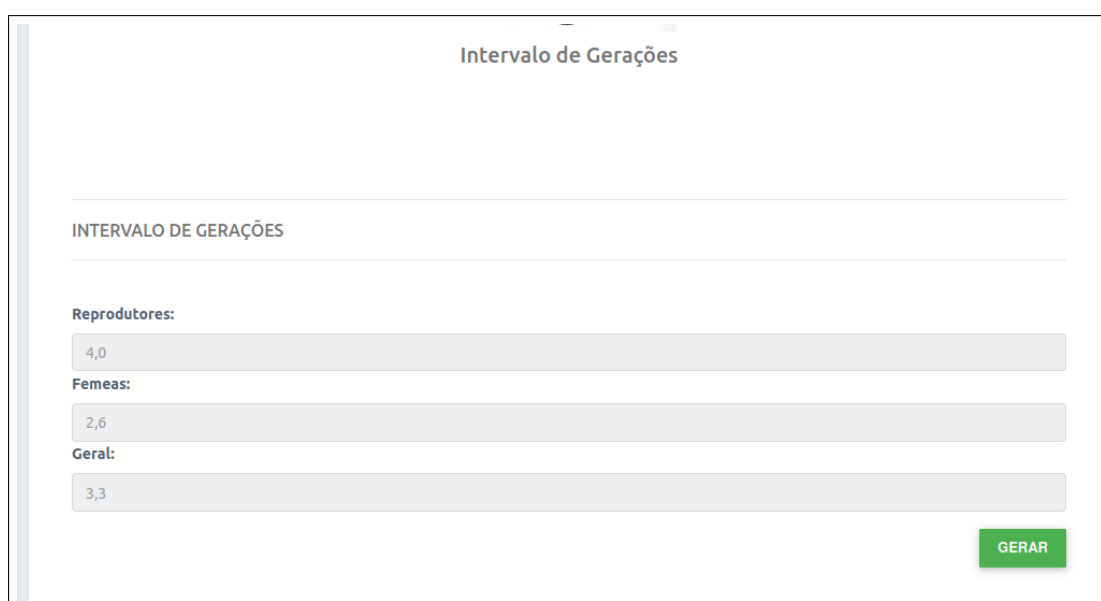


Figura 13 – Tela de mostragem do intervalo de gerações geral, de animais reprodutores e de animais matrizes.

6 Conclusão e Trabalhos Futuros

A criação de gado é uma atividade essencial para muitas comunidades ao redor do mundo, fornecendo alimentos e produtos para o consumo humano, além de ser uma importante fonte de renda. No entanto, a produção de gado pode ser desafiadora, especialmente quando se trata de alcançar resultados consistentes e de alta qualidade.

Este trabalho propôs a aplicação de conceitos de ciência de dados aplicados e integrados a um sistema de gerenciamento de rebanhos que usa os conceitos de endogamia e matriz de parentesco para fornecer dados para a tomada de decisão acerca de melhores cruzamentos com a finalidade de melhorar a qualidade genética do rebanho.

Ao utilizar essas dados, os criadores podem tomar decisões mais informadas sobre o manejo do rebanho e reduzir a necessidade de testes caros e demorados. Além disso, a utilização de sistemas de dados de ancestralidade e parentesco pode contribuir para o aumento da eficiência na produção de carne e leite, redução do impacto ambiental e melhoria da qualidade dos produtos finais.

Em conclusão, o sistema apresenta dados de ancestralidade e parentesco, além do cálculo de métricas úteis para a tomada de decisão quanto a cruzamentos e melhoria genética do gado, mostrando-se ferramenta útil para produtores e criadores que buscam melhorar a eficiência e a qualidade da produção.

A utilização de um sistema de dados de ancestralidade e parentesco pode ajudar os criadores a superar esses desafios, fornecendo informações importantes sobre a história genética do rebanho. Esses dados podem ser usados para identificar animais com características desejáveis e potenciais cruzamentos que podem resultar em uma prole saudável e produtiva.

Como trabalhos futuros, pretende-se ampliar a implementação com mais métricas de relevância para auxiliar a mão de obra qualificada no manejo do gado a tomar as melhores decisões de cruzamento do ponto de vista genético, bem como avaliar a implementação de métodos mais complexos de ciência de dados, como aprendizado de máquina e sugestões de cruzamentos.

Referências

- ABDALLA, E. Improving the reproductive performance of egyptian buffalo cows by changing the management system. *Animal reproduction science*, Elsevier, v. 75, n. 1-2, p. 1–8, 2003. Citado 2 vezes nas páginas 19 e 21.
- AHMAD, M. O.; MARKKULA, J.; OIVO, M. Kanban in software development: A systematic literature review. In: IEEE. *2013 39th Euromicro conference on software engineering and advanced applications*. [S.l.], 2013. p. 9–16. Citado na página 17.
- ANDRIETTA, L. T. *Uso de Machine Learning e dados genômicos para melhoria de características econômicas em bovinos de leite*. Tese (Doutorado) — Universidade de São Paulo, 2022. Citado 2 vezes nas páginas 20 e 21.
- BELLO, R.; ABUBAKAR, S. Development of a software package for cattle identification in nigeria. *Journal of Applied Sciences and Environmental Management*, v. 23, n. 10, p. 1825–1828, 2019. Citado na página 12.
- BOEG, J. Kanban em 10 passos. *Tradução de Leonardo Campos, Marcelo Costa, Lúcio Camilo, Rafael Buzon, Paulo Rebelo, Eric Fer, Ivo La Puma, Leonardo Galvão, Thiago Vespa, Manoel Pimentel e Daniel Wildt. C4Media*, p. 27, 2010. Citado na página 16.
- BOEHM, B.; ABTS, C.; CHULANI, S. Software development cost estimation approaches—a survey. *Annals of software engineering*, Springer, v. 10, n. 1-4, p. 177–205, 2000. Citado na página 16.
- BRIAND, L. C.; WIECZOREK, I. Resource estimation in software engineering. *Encyclopedia of software engineering*, Citeseer, v. 2, p. 1160–1196, 2002. Citado na página 16.
- BRUMATTI, R. C. et al. Desenvolvimento de índice de seleção em gado corte sob o enfoque de um modelo bioeconômico. *Archivos de zootechnia*, SciELO Espana, v. 60, n. 230, p. 205–213, 2011. Citado 2 vezes nas páginas 19 e 21.
- CHANDRA, V.; SEJIAN, V.; SHARMA, G. T. Strategies to improve livestock reproduction under the changing climate scenario. In: *Climate change impact on livestock: adaptation and mitigation*. [S.l.]: Springer, 2015. p. 425–439. Citado 2 vezes nas páginas 19 e 21.
- CORONA, E.; PANI, F. E. et al. A review of lean-kanban approaches in the software development. *WSEAS transactions on information science and applications*, WSEAS, v. 10, n. 1, p. 1–13, 2013. Citado na página 17.
- CROPS and livestock products. acessado em 06 de Março de 2023. <https://www.fao.org/faostat/en/#data/QCL>. Citado na página 12.
- DAUZON, S.; BENDORAITIS, A.; RAVINDRAN, A. *Django: web development with Python*. [S.l.]: Packt Publishing Ltd, 2016. Citado na página 17.
- EXPORTACOES. acessado em 06 de Março de 2023. <https://www.abiec.com.br/exportacoes/>. Citado na página 12.

- FARAH, M. M. Efeito da utilização de diferentes matrizes genômicas e parentesco na avaliação genética de bovinos de corte. Universidade Estadual Paulista (Unesp), 2014. Citado na página 15.
- FILHO, L. R. G. et al. Software to assess beef cattle body mass through the fuzzy body mass index. *Engenharia Agrícola*, SciELO Brasil, v. 36, p. 179–193, 2016. Citado na página 12.
- FONTOURA, D. et al. Optimizing mate selection: a genetic algorithm approach. In: IN: ICAR CONFERENCE, 43., 2019, PRAGUE. PROCEEDINGS... ROME: ICAR, 2019. [S.l.], 2019. Citado 3 vezes nas páginas 19, 20 e 21.
- FONTOURA, D. d. C. N. d. Uma solução de recomendações de acasalamentos baseada em algoritmos genéticos. Universidade Federal do Pampa, 2019. Citado 2 vezes nas páginas 14 e 20.
- FRY, J. P.; SIBLEY, E. H. Evolution of data-base management systems. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 8, n. 1, p. 7–42, 1976. Citado na página 15.
- GORDON, M.; HAREL, D. Steps towards scenario-based programming with a natural language interface. In: SPRINGER. *From Programs to Systems. The Systems perspective in Computing: ETAPS Workshop, FPS 2014, in Honor of Joseph Sifakis, Grenoble, France, April 6, 2014. Proceedings*. [S.l.], 2014. p. 129–144. Citado na página 12.
- HENDERSON, C. Use of relationships among sires to increase accuracy of sire evaluation. *Journal of Dairy Science*, Elsevier, v. 58, n. 11, p. 1731–1738, 1975. Citado na página 14.
- HENDERSON, C. R. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, JSTOR, p. 69–83, 1976. Citado na página 13.
- HOCQUETTE, J.-F. et al. Recent advances in cattle functional genomics and their application to beef quality. *Animal*, Cambridge University Press, v. 1, n. 1, p. 159–173, 2007. Citado na página 12.
- JULIUS, H. van der W. et al. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. In: BMC. [S.l.], 2012. Citado na página 15.
- KESHAVARZI, H. et al. Machine learning algorithms, bull genetic information, and imbalanced datasets used in abortion incidence prediction models for iranian holstein dairy cattle. *Preventive veterinary medicine*, Elsevier, v. 175, p. 104869, 2020. Citado 2 vezes nas páginas 19 e 21.
- KIROVSKA, N.; KOCESKI, S. Usage of kanban methodology at software development teams. *Journal of applied economics and business*, Education and Novel Technology Research Association, v. 3, n. 3, p. 25–34, 2015. Citado na página 17.
- KUDO, A. A method for calculating the inbreeding coefficient. *American journal of human genetics*, Elsevier, v. 14, n. 4, p. 426, 1962. Citado na página 12.

- LINDSTROM, L.; JEFFRIES, R. Extreme programming and agile software development methodologies. In: *IS management handbook*. [S.l.]: Auerbach Publications, 2003. p. 531–550. Citado na página 16.
- LIU, Q.; LANARI, M.; SCHAEFER, D. A review of dietary vitamin e supplementation for improvement of beef quality. *Journal of animal science*, Oxford University Press, v. 73, n. 10, p. 3131–3140, 1995. Citado na página 12.
- LOBO, N. P. Melhoria genética em bovino de corte. Pontifícia Universidade Católica de Goiás, 2021. Citado 2 vezes nas páginas 19 e 21.
- MACEDO, A. A. et al. Influence of endogamy and mitochondrial dna on immunological parameters in cattle. *BMC veterinary research*, Springer, v. 10, p. 1–10, 2014. Citado na página 12.
- MARTINS, E. *Uso de modelos mistos na avaliação genética animal*. [S.l.]: UFV, 1997. Citado 3 vezes nas páginas 23, 24 e 28.
- MOTTA, B. E. T. da; FERREIRA, A. P. L.; YOKOO, M. J.-I. Brangusselection: um algoritmo para seleção ótima de acasalamentos com índice de seleção customizável. In: SBC. *Anais do XIII Congresso Brasileiro de Agroinformática*. [S.l.], 2021. p. 35–43. Citado 2 vezes nas páginas 19 e 21.
- OUHBI, S.; POMBO, N. Software engineering education: Challenges and perspectives. In: IEEE. *2020 IEEE Global Engineering Education Conference (EDUCON)*. [S.l.], 2020. p. 202–209. Citado na página 16.
- PEREIRA, J. C. C. *Melhoria genética aplicado à produção animal*. [S.l.]: FEPMVZ Editora 4. ed. Belo Horizonte, 2004. Citado na página 13.
- PONTES, R. E. d. S. et al. Contratação do desenvolvimento ágil de software na administração pública federal: riscos e ações mitigadoras. Universidade Católica de Brasília, 2014. Citado na página 16.
- POPULACAO mundial chegara a 8 bilhoes em novembro de 2022. acessado em 06 de Março de 2023. <https://brasil.un.org/pt-br/189756-populacao-mundial-chegara-8-bilhoes-em-novembro-de-2022>. Citado na página 12.
- POWELL, J. E.; VISSCHER, P. M.; GODDARD, M. E. Reconciling the analysis of ibd and ibs in complex trait studies. *Nature Reviews Genetics*, Nature Publishing Group UK London, v. 11, n. 11, p. 800–805, 2010. Citado na página 15.
- QUEIROZ, S. A. d.; ALBUQUERQUE, L. G. d.; LANZONI, N. A. Efeito da endogamia sobre características de crescimento de bovinos da raça gir no brasil. *Revista Brasileira de Zootecnia*, SciELO Brasil, v. 29, p. 1014–1019, 2000. Citado na página 14.
- RAMAKRISHNAN, R.; GEHRKE, J.; GEHRKE, J. *Database management systems*. [S.l.]: McGraw-Hill New York, 2003. Citado na página 15.
- ROSA, A. d. N. et al. Melhoria genética aplicado em gado de corte: Programa geneplus-embrapa. Brasília, DF: Embrapa; Campo Grande, MS: Embrapa Gado de Corte, 2013., 2013. Citado na página 14.

- SILVA, L. A. et al. Ciência de dados educacionais: definições e convergências entre as áreas de pesquisa. In: *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*. [S.l.: s.n.], 2017. v. 6, n. 1, p. 764. Citado na página [15](#).
- STOPA, G. R.; RACHID, C. L. Scrum: Metodologia ágil como ferramenta de gerenciamento de projetos. *CES Revista*, v. 33, n. 1, p. 302–323, 2019. Citado na página [16](#).
- UPPERMAN, L. R. et al. Management of lethal recessive alleles in beef cattle through the use of mate selection software. *Genetics Selection Evolution*, BioMed Central, v. 51, n. 1, p. 1–16, 2019. Citado na página [12](#).
- VAMSI, K. M. et al. Visualization of real world enterprise data using python django framework. In: IOP PUBLISHING. *IOP Conference Series: Materials Science and Engineering*. [S.l.], 2021. v. 1042, n. 1, p. 012019. Citado na página [18](#).



TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DIGITAL NA BIBLIOTECA “JOSÉ ALBANO DE MACEDO”

Identificação do Tipo de Documento

- () Tese
() Dissertação
(X) Monografia
() Artigo

Eu, Samuel Pedro Beserra Dantas Lélis, autorizo com base na Lei Federal nº 9.610 de 19 de Fevereiro de 1998 e na Lei nº 10.973 de 02 de dezembro de 2004, a biblioteca da Universidade Federal do Piauí a divulgar, gratuitamente, sem ressarcimento de direitos autorais, o texto integral da publicação **“Aplicação de ciência de dados para a tomada de decisão em acasalamento de bovinos e melhoria genética”** de minha autoria, em formato PDF, para fins de leitura e/ou impressão, pela internet a título de divulgação da produção científica gerada pela Universidade.

Picos-PI 24 de Março de 2023.

Samuel Pedro B. D. Lélis

Assinatura

Assinatura